Part II

Hints for Advanced Placement Statistics Courses

AP Features of the Text

Understandable Statistics is particularly well-suited for Advanced Placement statistics courses. The standard topics are carefully presented, the Guided Exercises invite students to be active learners, and there is an emphasis on interpreting statistical results in real-world contexts.

Technology support for the TI-83 Plus, MINITAB, and Microsoft Excel are included in the text. Two technology supplements, the *Technology Guide* and the *Excel Guide,* give detailed instructions for the use of various technologies and contain supplemental learning activities to help students use technology to explore statistics.

The seventh edition also incorporates a number of features to make the text even more useful for AP courses:

- The normal distribution table has been changed to the cumulative, left-tailed style provided during the AP exam.
- Section 1.3, Introduction to Experimental Design, is a new section on issues related to the planning of statistical studies.
- Dotplots are discussed in Problems 17, 18, and 19 of Section 2.2.
- Section 5.1, Introduction to Random Variables and Probability Distributions, now includes the mean and standard deviation of a linear transformation of a random variable. The section also includes the mean and standard deviation for linear combinations of independent random variables. (The mean and standard deviation for a linear combination of two random variables are discussed in Problem 16 of Section 10.3.)
- A discussion of normal quantile plots is included in the Using Technology section of Chapter 6.
- Section 7.3, Sampling Distributions for Proportions, is a new section. The topic of this section is also addressed in sections on confidence intervals and the testing of proportions.
- Residual plots for linear regression are discussed in Problems 17 and 18 of Section 10.2.
- Section 10.4, Inferences Concerning Regression Parameters, has been revised. Tests of the correlation coefficient are now done with Student's t distribution rather than a special table for r. Additional topics include testing and confidence intervals for the slope of the least-squares line.
- The Instructor's Annotated Edition now includes margin answers for both odd and even problems. In addition, teaching hints and comments continue to be provided within each section.
- Headers for each example and Guided Exercise in the text quickly identify the statistical analysis being demonstrated. Headers for each of the section and chapter problems identify the field of study from which the application is drawn.

In addition, the Test Item File supplement contains a special section of questions geared to preparing students for the AP Exam. It contains 48 multiple-choice questions and 12 free-response questions. The first 11 free-response questions are similar to the types of questions students will see on the Part A of the free-response section of the AP Exam while the last question is similar to the longer more in-depth free-response question the students will see in Part B.

Information about the AP Statistics Exam

OUTLINE OF MAJOR TOPICS COVERED IN THE AP STATISTICS EXAM

I. Exploring Data: Observing patterns and departures from patterns
Use of graphical and numerical techniques to study patterns and departure from patterns, with an emphasis on interpreting the information from graphical displays and numerical summaries.

A. Single-variable data
Students should be able to construct and interpret graphs such as histograms, stem-and-leaf plots, and box-and-whisker plots. They must also be well-versed in different ways of describing the shape, center, and variability of a distribution of data values and the position of individual values relative to the set.

- A distribution's shape may variously be described as (among other things) either symmetric or asymmetric, approximately normal, or uniform.
- Summary statistics such as the mean, median, and mode give measures of the center.
- Values such as the variance, standard deviation, range, and interquartile range describe the variability.
- Quartiles, percentiles, standardized scores (z-scores) and outlier values give information about the position of an individual data value relative to the overall distribution.

Students should also understand the effect that changing units has on the different summary measures.

B. Data pairs
Students should be able to use scatter diagrams to detect patterns in data pairs and to find the equation of the least-squares regression line. They should know how to use the plot of residuals to judge the appropriateness of a linear model, and how to use the correlation coefficient to judge a linear model's strength. Students should be able to identify outliers and influential points. They should also understand the relation of causation and correlation.

II. Planning a Study: Deciding what and how to measure
In open-ended questions, students need to demonstrate that they understand how to gather data according to a well-developed plan. Methods of data collection such as census, sample survey using simple random samples, experiment, and observational study may be used. Students should demonstrate an awareness of sampling error, sources of bias, and use of stratification to reduce variation. Experimental designs, such as completely randomized design for two treatments and blocking designs, might be required.

III. Anticipating patterns: Producing models using probability and simulation
Under this topic, students use probability as a tool for anticipating what the distribution of data should look like for a given model. Particular topics include

- rules of probability
- law of large numbers
- independence of events
- mean and standard deviation of a random variable
- use of binomial distributions
- normal distributions and sampling distributions of a sample proportion
- sample mean
- difference of two independent sample proportions
- difference between two independent sample means

IV. Statistical inference: Confirming models
Students should know how to select appropriate models for statistical inference.

A. Confidence intervals
- their meaning
- large-sample confidence interval for a proportion
- large-sample confidence interval for a mean
- large-sample confidence interval for a difference between two proportions
- large-sample confidence interval for a difference between two means (unpaired and paired)

B. Tests of significance
- structure and logic, including null hypothesis, alternate hypothesis, P values, one- and two-tailed tests
- large-sample tests for a mean and a difference between two means (paired and unpaired)
- large-sample tests for a proportion and a difference between two proportions
- chi-square test for goodness of fit, homogeneity of proportions, and independence

C. Special case of normally distributed data
- Student's t distribution
- single-sample t procedures
- two-sample t procedures (independent and matched pairs)
- inference for slope of least-squares line

TYPES OF QUESTIONS

The AP test has two parts, the first of which is multiple-choice. The second part is a free-response section requiring the student to answer open-ended questions and to complete an investigative task involving more extended reasoning. The two sections are given equal weight in determining the grade for the examination.

DURATION AND GRADING OF THE EXAM

The test is three hours long. The exam consists of two sections, each of which will last 90 minutes and will each account for 50% of the total exam grade: a multiple-choice section that consists of 40 problems, and a second section of six free-response problems. Students will be instructed before the multiple-choice section that they most likely will not have time to answer every multiple-choice problem, and that they do not benefit from guessing because 1/4 of the number of incorrect responses on the multiple-choice section will be deducted from the number of correct responses. The free-response section consists of two parts. Part A will be made up of 5 free-response problems for which they will be allowed 65 minutes. Part A will account for 75% of the free-response section. Part B consists of a longer more in-depth free-response problem. Students will be allowed 25 minutes for this part which will count as 25% of the free-response score.

CALCULATOR

Each student is expected to use a graphing calculator with statistics capabilities during the AP exam. *Understandable Statistics* uses the versatile and widely available TI-83 Plus.

FORMULAS AND TABLES

Formulas and tables are provided for students taking the AP Statistics Examination. The format of the formulas and some of the tables is slightly different from those in the text. Note that the normal distribution table for the AP exam gives areas in the *left tail* of the distribution. (Some references give the area from 0 to z instead; *Understandable Statistics* contains both types of tables.) The table for Student's t distribution provides critical values for areas in the right tail of the distribution, and gives critical values for different confidence levels.

The following formulas are provided during the AP exam.

Descriptive Statistics

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}}$$

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b_1 = r\frac{s_y}{s_x}$$

$$s_{b_1} = \frac{\sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}}{\sqrt{\sum(x_i - \bar{x})^2}}$$

Probability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$E(X) = \mu_x = \sum x_i p_i$$

$$Var(X) = \sigma_x^2 = \sum(x_i - \mu_x)^2 p_i$$

If X has a binomial distribution with parameters n and p, then:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mu_x = np$$

$$\sigma_x = \sqrt{np(1-p)}$$

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

If $\bar{x}$ is the mean of a random sample of size n from an infinite population with mean μ and standard deviation σ, then:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

Confidence interval: statistic $\pm$ (critical value) $\cdot$ (standard deviation of statistic)

Single-Sample	
Statistic	Standard Deviation of Statistic
Sample Mean	$\dfrac{\sigma}{\sqrt{n}}$
Sample Proportion	$\sqrt{\dfrac{p(1-p)}{n}}$

Two-Sample	
Statistic	Standard Deviation of Statistic
Difference of sample means	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ Special case when $\sigma_1 = \sigma_2$ $\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$
Difference of sample proportions	$\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ Special case when $p_1 = p_2$ $\sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

$$\text{Chi-square test statistic} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

DIFFERENCES IN CONVENTION AND NOTATION BETWEEN *UNDERSTANDABLE STATISTICS* AND THE AP EXAM

Notation and Terminology Correlation Chart	
Understanding Statistics	**AP Statistics Exam**
Box-and-whisker plot	Box Plot
Alternate hypothesis: H_1	Alternate hypothesis: H_a
Smallest level of significance at which to reject the null hypothesis: P-value	Smallest level of significance at which to reject the null hypothesis: p-value
Conditional probability: $P(A \text{ given } B)$	Conditional probability: $P(A \mid B)$
Ogive	Cumulative graph

SOME COMMON STUDENT MISTAKES ON THE AP EXAM

Students taking the AP exam often make the following mistakes:

- Failing to follow through with the answer to a question. For example, students asked to compare two data sets may describe each set but fail to finish with a statement that compares the one to the other.
- Getting mixed up about what the numbers in data sets mean. For example, students will sometimes treat one-variable data in a frequency table as if it were paired data, and construct a scatter diagram when a histogram is called for.
- Misreading the computer printouts that appear as graphics on the test. For example. a student may take a number to represent the slope of the regression line when in fact this number represents the constant term.
- Getting confused by the coding for representing data values of variables. For example, a student may forget that if x represents years after 1900, then the year 1995 is represented by $x = 5$. not $x = 1995$.

MORE INFORMATION AND SAMPLE QUESTIONS

Advanced Placement Program Course Description - Statistics, published by The College Board, contains sample questions, formulas, tables, and a description of the AP Statistics course. This publication is available from the College Board.

The College Board
45 Columbus Avenue
New York, NY 10023-6992
Phone (212) 713-8000

Additionally, old AP Statistics tests are available. The College Board releases the free-response part of the exam every year and the multiple-choice part once every five years. The free-response sections and sample solutions are available for free on the College Board website at www.collegeboard.org/ap/statistics. Copies of the multiple-choice part of the exam can be purchased from the College Board by calling or by visiting the "Shop" section on the College Board website.

Hints for Helping Students Succeed on the AP Statistics Exam

DEVELOP THEIR COMMUNICATION SKILLS

On the free-response portion of the AP Statistics Exam, graders will be looking for students to show that they understand the statistical knowledge and that they are able to communicate that knowledge clearly. Throughout the course, give students lots of opportunities to explain their reasoning. Encourage them to show their thinking in a variety of ways. They should use diagrams where appropriate and words to explain the reasons they choose to use certain techniques. Encourage them to write as if they are explaining their ideas to someone who has not yet mastered statistics.

DEVELOP THEIR GRAPHING CALCULATOR SKILLS

Give students lots of opportunities to develop their skills using a graphing calculator to do statistics. The more familiar they are with the capabilities of their calculator and how to use them, the more time they will have during the AP Statistics Exam to devote to their solution.

DEVELOP THEIR SKILLS USING STATISTICAL TABLES AND FORMULAS

Be sure that students can quickly use the statistical tables and apply the formulas that are provided on the AP Statistics Exam.

EMPHASIZE CONCEPTUAL UNDERSTANDING

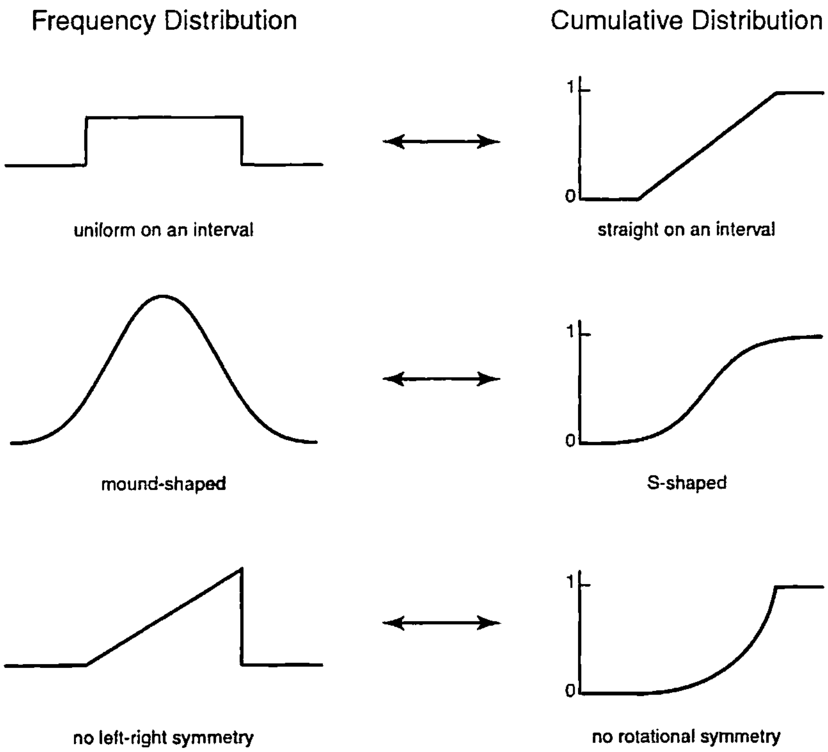
Displaying Data

Be sure students know the appropriate uses of the various graphs. They should use

- Bar graphs when comparing data sets;
- Circle graphs for displaying how a total is dispersed into several categories;
- Time plots to display how data change over time;
- Histograms or frequency polygons to display relative frequencies of grouped data;
- Stem-and-leaf displays for displaying grouped data in a way that does not lose the detail of the original raw data.

Drawing and Using Ogives (Section 2.2)

Be sure students understand the relationship between the shapes of the frequency distribution curve and the cumulative distribution curve. The one can be used to infer the other, as illustrated on the next page.



Quartiles

The text mentions, in passing, that there are different conventions for calculating the first and third quartiles of a data set. The convention in the text is that the first quartile is the median of the data set's lower half, *not including* the data set's overall median (the second quartile). Correspondingly, the third quartile is the median of the upper half, not including the overall median. The usual alternate method is to include the overall median in both the upper and the lower half of the data set. Then it plays a role in the calculation of both the first and the third quartile.

To illustrate how it matters which method you choose, use both methods to calculate the first and third quartiles of the data set below, where $N = 15$, an odd number.

Test scores														
59	63	68	73	74	76	81	84	85	88	91	95	95	97	98
							median							

When the median is left out of the lower and upper halves of the data, the first quartile is the fourth of the first seven values, namely 73, and the third quartile is the fourth of the last seven values, namely 95. When the median is included in the lower and upper halves of the data, the first quartile is halfway between the fourth and fifth of the first eight values, which makes it 73.5. And the third quartile is halfway between the fourth and fifth of the last eight values, which makes it 93.

The two methods always give the same results when N, the size of the data set, is an even number, because then the issue of whether or not to include the median in each half does not arise.

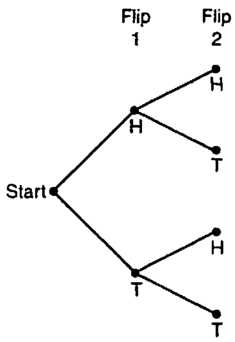
Ways to Think about Probability

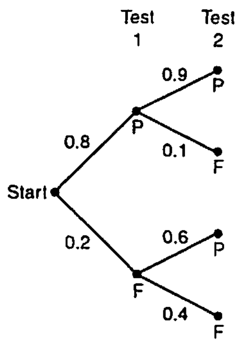
As the text describes in Section 4.1, there are several methods for assigning a probability to an event:

- Probability based on intuition is often called *subjective* probability. Thus understood, probability is a numerical measure of a person's confidence about some event. Subjective probability is assumed to be reflected in a person's decisions: the higher an event's probability, the more the person would be willing to bet on its occurring.
- Probability based on relative frequency often called *experimental* probability, because relative frequency is calculated from an observed history of experiment outcomes. It can also be called *observed* probability.
- Probability based on equally likely outcomes is often called *theoretical* probability, because it is ultimately derived from a theoretical model of the experiment's structure. The experiment may be conducted only once, or not at all, and need not be repeatable.

These three ways of treating probability are compatible and complementary. For a reasonable, well-informed person, the subjective probability of an event should match the theoretical probability, and the theoretical probability in turn predicts the observed probability as the experiment is repeated many times.

For purposes of the AP exam, the primary understanding of probability is in terms of relative frequency. Test questions involving probability typically refer to past results of some experiment (such as clinical trials of a medication). However, students must be able to manipulate the information in the question appropriately. They should be able to use tree diagrams not only on experiments with equally likely outcomes, as in Section 4.3 (see the figure below on the left) but also on experiments where only probabilities of various events are given (see the figure on the right).





Probability Statements with > and ≥: Discrete versus Continuous Variables

When x is a discrete variable, there is a real difference between, say, $P(x > 5)$, the probability that x is greater than 5, and $P(x \geq 5)$, the probability that x is greater than or equal to 5. The difference comes from the fact that $P(x = 5)$, the probability that x equals 5, is in general greater than zero. By contrast, when x is a continuous variable, $P(x = 5)$ is essentially zero: the probability that the variable takes exactly the value 5 is vanishingly small. This means that there is no important difference between $P(x > 5)$ and $P(x \geq 5)$ when x is a continuous variable.

Binomial Probabilities (Section 5.2)

Students should be able to show that $p(1 - p) = pq$ has its maxim value at $p = 0.5$. There are at least three ways to demonstrate this: graphically, algebraically, and using calculus.

Graphical and algebraic methods

See the Teaching Tips for Chapter 5.

Calculus-based method

Advanced Placement students have probably had (or are taking) calculus, including tests for local extrema. For a function with continuous first and second derivatives, at an extremum the first derivative equals zero and the second derivative is either positive (at a minimum) or negative (at a maximum).

The first derivative of $f(p) = pq = p(1-p)$ is given by

$$\begin{aligned} f'(p) &= \frac{d}{dp}[p(1-p)] \\ &= \frac{d}{dp}[-p^2+p] \\ &= -2p+1 \end{aligned}$$

Solve $f'(p) = 0: -2p+1 = 0$

$$p = \frac{1}{2}$$

Now find $f''\left(\frac{1}{2}\right)$: $f''(p) = \frac{d}{dp}[f'(p)]$

$$\begin{aligned} &= \frac{d}{dp}(-2p+1) \\ &= -2 \end{aligned}$$

So $f''\left(\frac{1}{2}\right) = -2$.

Since the second derivative is negative when the first derivative equals zero, $f(p)$ has a maximum at $p = 1/2$.

Normal Approximation to the Binomial Distribution (Section 6.4)

As a test of the suitability of the normal approximation to the binomial distribution, *Understandable Statistics* uses the test $np > 5$, $nq > 5$. The AP exam, however, expects students to use the test $np > 10$, $nq > 10$. This test is based on recent studies indicating that the >5 test is not adequate for very small p or q.

Continuity Correlation in Sampling Distributions for Proportions (Section 7.3)

Emphasize the difference in the continuity correction that must be taken into account in a sampling distribution for proportions and the continuity correction for a normal distribution used to approximate the probability distribution of the discrete random variable in a binomial probability experiment. That is, instead of subtracting 0.5 from the left endpoint and adding 0.5 to the right endpoint of an interval involved in a normal distribution approximating a binomial probability distribution, $0.5/n$ must be subtracted from the left endpoint and $0.5/n$ added to the right endpoint of such an interval, where n is the number of trials, when a normal distribution is used to approximate a sampling distribution for proportions.

Understanding Confidence Intervals (Section 8.1)

Students need to understand that the method for creating confidence intervals is not a procedure that comes up with a statement having a specified probability (like 90%, or 95%) of being true. Instead, it is a procedure that has a specified probability of coming up with a true statement. These are two different things.

To understand the difference, consider a simple task: guessing the color of a randomly-drawn marble. Suppose a sack contains some red marbles and some blue marbles. And suppose we have a friend who will reach in, draw out a marble, and announce its color while we have our backs turned. The friend can be counted on to announce the correct color *exactly 90% of the time* (that is, with a probability of 90%) and the wrong color the other 10% of the time. So if the marble drawn is blue, the friend will say "blue" 9 times out of 10 and "red" the remaining time; and conversely for a red marble. This is like creating an 0.90 confidence interval.

The friend reaches in, pulls out a marble, and announces, "blue." Does this mean that we are 90% sure the friend is holding a blue marble? Not necessarily. Suppose we think that the bag contains three red marbles and two blue ones. Then we expect the friend to draw a red marble 3/5 of the time and announce "blue" 10% of those times, or 3/50 of all draws. And we expect the friend to draw a blue marble 2/5 of the time and announce "blue" 90% of those times. or 18/50 of all draws. This means that the ratio of true "blue" announcements to false ones is 18 to 3, or 6 to 1. And thus we attach a probability of 6/7 = 85.7%. not 90%, to our friend's announcement that the marble drawn is blue. even though we believe our friend to be telling the truth 90% of the time. Simply put, our initial belief that there are more red marbles than blue ones pulls our confidence in a "blue" announcement downward.

Now, if we believe that there are an *equal* number of red and blue marbles in the bag, then, as it turns out, we will attach 90% probability to "blue" announcements and to "red" announcements as well. But *this is a special case.* In general, the probabilities we attach to each of our friend's statements will be different from the frequency with which we think he is telling the truth.

For the same reasons, when we create a 0.90 confidence interval, we will in general not be 90% sure that the interval contains the mean. It could *happen* to turn out that we were 90% sure, but this will depend on what ideas we had about the population mean going in. Suppose we were fairly sure, to start with, that the population mean lay somewhere between 10 and 20, and suppose we then took a sample that led to the construction of a 0.90 confidence interval which ran from 31 to 46. We would *not* conclude, with 90% certainty, that the mean lay between 31 and 46. Instead, we would have a probability lower than that, because previously we thought the mean was outside that range. At the same time, we would be much more ready to believe that the mean lay between 31 and 46 than we were before, because, after all, a procedure with a 90% success rate produced that prediction.

To repeat: the procedure for finding a confidence interval of confidence level c does not, in general, produce a statement (about the value of a population mean) that has a probability c of being true. This issue is discussed somewhat more fully in the Teaching Tips for Chapter 8.

Hypothesis testing and *P* values (Sections 9.1–9.3)

Hypothesis Tests

The hypothesis tests in these sections test the value of a population mean, μ, against some specified value, denoted by k. Be sure students understand when to use each test.

Z-tests are appropriate for testing the null hypothesis $H_0: \mu = k$ against one of the three alternative hypotheses: $H_1: \mu > k$, $H_1: \mu < k$, or $H_1: \mu \neq k$ when (1) the data in the sample are known to be from a normal distribution (in which case, the sample may be any size), or when (2) the data distribution is unknown or the data are believed to be from a non-normal distribution, but the sample size is large $(n > 30)$. The z-test requires the population standard deviation σ to be known, but if the sample size n is large, the sample standard deviation, s, is assumed to be close to σ, and can be used instead.

For an upper- or right-tailed test, the alternative hypothesis is $H_1: \mu > k$, where k is the hypothesized value of the population mean and the sample mean $\bar{x}$ is greater than k. (If $\bar{x}$ is less than or equal to k, you do not use this test.) This test calculates the sample standard deviation s from the sample data and substitutes s for σ in the formula $z_{calculated} = \frac{\bar{x}-k}{\sigma/\sqrt{n}}$. The P value is then calculated as the probability of getting a z-value as large or larger than the one observed (that is, $z_{calculated}$) if the null hypothesis is true; small P values indicate that the null hypothesis should be rejected.

For a lower- or left-tailed test, the alternative hypothesis is $H_1: \mu < k$, where k is the hypothesized value of the population mean and the sample mean $\bar{x}$ is less than k. (If $\bar{x}$ is greater than or equal to k, you do not use this test.) This test calculates the sample standard deviation s from the sample data and substitutes s for σ in the formula $z_{calculated} = \frac{\bar{x}-k}{\sigma/\sqrt{n}}$. The P value is then calculated as the probability of getting a z-value as small or smaller than the one observed (that is, $z_{calculated}$) if the null hypothesis is true; small P values indicate that the null hypothesis should be rejected.

For a two-tailed test, the alternative hypothesis is H_1: $\mu \neq k$, where k is the hypothesized value of the population mean. This test calculates the sample standard deviation s from the sample data and substitutes s for σ in the formula $z_{calculated} = \frac{\bar{x}-k}{\sigma/\sqrt{n}}$. The P value is then calculated as the probability of getting a z-value as far or further away from zero as the one observed (that is, $z_{calculated}$) if the null hypothesis is true. This P value is always double the value found using the appropriate one-tailed test (right-tailed for $\mu > k$, left-tailed for $\mu < k$). Small P values indicate that the null hypothesis should be rejected.

P Values

Suppose you and a friend were tossing a coin 100 times, and if there are more than 50 heads, you buy ice cream for both of you; if there are fewer than 50 heads, your friend buys, and if there are exactly 50 heads, each person buys his or her own ice cream. You and your friend plan to toss the coin and have ice cream every day after school. You reason, correctly, that over the school year, you will buy about half the time, and your friend will buy about half the time IF the coin you are tossing is fair. However, after a week or so, you notice that there seem to be lots of heads and that you seem to be doing most of the buying. You also notice that your friend always seems to have a coin handy, and most of the time you use his coin for the coin toss. Where do you draw the line between "I'm just unlucky" and "That jerk is cheating!"? At 55 heads out of 100 tosses? At 60 heads? At 75 heads? At 90 heads?

The *P* value is helping to draw the line for you. It tells you how likely you are to see, say, 60 or more heads in 100 tosses of a fair coin. If this probability is, in your opinion, too small, you know whom *not* to call your friend!

Many problems specify testing at level α; in these problems, if the P value is less than α, reject H_0. Frequently-used α are 0.10, 0.05, and 0.01; however, there is nothing sacred or special about these values. If you want to use an alpha of 0.0827 as your cutoff point for "too small," you can. The numbers in the tables for z values and t values were originally obtained by hand, using calculus and old-fashioned mechanical calculators. The amount of work that went into calculating the values cannot be overstated, and when these tables were first printed in the 1920s, 1930s, and 1940s, it made life much easier for practicing statisticians. If you wanted to know which value of Student's t with 14 degrees of freedom cut off an area of 0.04135 (instead of 0.05 or 0.025, say) in the upper tail, you either said that the t-value was between the tabled values 1.761 and 2.145, or you calculated it yourself. Before there were powerful and readily accessible computers, most people said tail areas of 0.05 or 0.025 were good enough for them. And that is why 0.10, 0.05, 0.01, etc., are so common—we have always used these numbers. Even today, at the bottom of z, t, etc., tables in textbooks, you can find statements such as "reprinted with the permission of the *Biometrika* trustees," a reference to the fact that these tables are copies of the tables that originally appeared in the journal *Biometrika* many years ago.

Understanding the Results of a Hypothesis Test

Just as students may get confused about the probability associated with a confidence interval, they may get confused about the probability associated with a hypothesis test. The *P* value is *not* the probability, in light of the data, of the null hypothesis. Instead, the *P* value is the probability that the data would turn out the way it did, assuming the null hypothesis to be true.

To illustrate: consider two coins in a sack, one fair and one two-headed. One of these coins is pulled out at random and flipped. It comes up heads. Let us take, as our null hypothesis, the statement "The flipped coin was the fair one." The probability of the outcome, given the null hypothesis, is 1/2, because a fair coin will come up heads half the time. This probability is in fact the *P* value of the outcome. On the other hand, the probability that the null hypothesis is true, given the evidence, is 1/3, since out of all the heads outcomes one will see in many such trials, 1/3 are from the fair coin.

The meaning of the probability associated with a hypothesis test is more fully discussed in the Teaching Tips for Chapter 9.